Journal of Cognitive Psychology

Perceptual and categorical judgements of colour similarity

Jules Davidoff, Julie Goldstein, Ian Tharp, Elley Wakui & Joel Fagot

Psychology Department, Goldsmiths, University of London, London, UK

Laboratoire de Psychologie Cognitive, University of Provence, Marseille, France

Version of record first published: 24 Jul 2012

To cite this article: Jules Davidoff, Julie Goldstein, Ian Tharp, Elley Wakui & Joel Fagot (2012): Perceptual and categorical judgements of colour similarity, Journal of Cognitive Psychology, DOI:10.1080/20445911.2012.706603

To link to this article: http://dx.doi.org/10.1080/20445911.2012.706603

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Perceptual and categorical judgements of colour similarity

Jules Davidoff1, Julie Goldstein1, Ian Tharp1, Elley Wakui1, and Joel Fagot2

1Psychology Department, Goldsmiths, University of London, London, UK
2Laboratoire de Psychologie Cognitive, University of Provence, Marseille, France

Consideration is given to the tasks that make judgements of colour similarity based on perceptual similarity rather than categorical similarity. Irrespective of whether colour categories are taken to be universal (Berlin & Kay, 1969) or language induced (Davidoff, Davies, & Roberson, 1999), it is widely assumed that colour boundaries, and hence categorical similarity, would be used when categorising colours. However, we argue that categorical similarity is more reliably used in implicit than in explicit categorisation. Thus, in Experiment 1, we found that category boundaries may be overridden in the explicit task of matching-to-sample: There was a similar strong tendency to ignore colour boundaries and to divide the range of coloured stimuli into two equal groups in both Westerners and in a remote population (Himba). In Experiment 2, we showed that a distinctive stimulus (focal colour) in the range affected the equal division in a matching-to-sample task. Experiment 3 tested the stability of a category boundary in an implicit task (visual search) that assessed categorical perception; only for this task was categorisation largely immune to range effects and largely based on categorical similarity. It is concluded that, even after colour categories are acquired, perceptual rather than categorical similarity is commonly used in judgements of colour similarity.

Keywords: Colour similarity; Categorical similarity; Perceptual similarity; Colour naming.

Two ways have been proposed for judging colour similarity (Goldstein, 1948; Kay & Kempton, 1984; Roberson, Davidoff, & Braisby, 1999; Roberson, Hanley, & Pak, 2009); ways that we will call perceptual similarity and categorical similarity. Perceptual similarity as defined here is dictated by the physiology of colour vision and will be identical for observers with normal visual sensory function. Judgements by perceptual similarity produce some discontinuities in colour space but, by themselves, (see Webster & Kay, 2011) would not produce even the basic colour categories of Berlin and Kay (1969). We cannot rule out the possibility that primary colour prototypes (foci) have a physiological basis (Kay & Regier, 2006; Philipona & O’Regan, 2006) but it is also clear that what counts as a focal colour for one language does not do so in other (Roberson, Davidoff, Davies, & Shapiro, 2004). A more complex account must therefore be given for the wide range of colour names (categories) and hence categorical similarity judgements used by speakers of the world’s languages.

One recent approach to explain colour categories, popular for computer simulations, is to argue that categories optimize the communication benefits of a solution for any division of colour space (Komarova, Jameson, & Narens, 2007;
The underlying principle in these models can be traced back to Jameson and D’Andrade (1997) and in turn to ideas of Garner (1974) that an optimal solution would minimise within-category similarity and maximise between-category similarity. Some formulations for the optimal partition have favoured a solution based on the Berlin and Kay basic colours (Dowman, 2007; Regier et al., 2007), but others have only weakly (Griffin, 2006) or have no special place for them (Komarova et al., 2007). However, these partitions are only going to explain part of the data on colour categories. Their first weakness is that they have no mechanism whereby observers may switch from categorical to perceptual judgements; a switch that is argued to rely on the use, or not, of language networks (Davidoff & Roberson, 2004; Kay & Kempton, 1984; Roberson et al., 1999). A second weakness common to these models (Komarova et al., 2007; Regier et al., 2007) is that they do not explain how different boundaries arise between languages. For example, two 5-term colour languages from very different cultures and environments studied in Roberson, Davidoff, Davies, and Shapiro (2005) appear to divide colour space in similar ways but the foci and boundaries were specific to each language with specific behavioural consequences for learning and memory.

An alternative approach to colour categories is to link them to colour names. More than that, the claim is that the linking to a name may cause a change in perception—colours with different names (from different categories) look more different, and colours from the same category look more similar, than would be predicted from our physiology (Kay & Kempton, 1984). Thus, categorical colour similarity, as for the categorisation of continua in the auditory domain, implies the property of Categorical Perception (CP; Harnad, 1987). In practice, CP is assessed by performance with a target being more accurate/faster when foils are from different categories than when they are from the same category despite all foils being at an equal separation from the target as judged by perceptual similarity.

The cross-category advantage (CP) was linked to language by Davidoff et al. (1999); indeed, subsequent neuroimaging research has shown activity in language areas during colour similarity judgements (Siok et al., 2009; Tan et al., 2008). One particularly strong line of supporting evidence is that a concurrent verbal task removes CP (Gilbert, Regier, Kay, & Ivry, 2006; Roberson & Davidoff, 2000; Winawer, Witthoft, Frank, Wu, & Boroditsky, 2007). Importantly, the interference task to remove CP does not require rehearsing colour names (Roberson & Davidoff, 2000). Even though colour CP derives from a language’s colour terms (e.g., Roberson et al., 2004), the change in appearance of colours has become linked to a neural network that involves frontal areas, language areas but not those that when damaged result in colour anomaia, and the visual cortex (Tan et al., 2008).

Not all researchers go along with a CP account of categorical similarity. One early account regarded colour names as irrelevant (Rosch Heider & Olivier, 1972). They proposed that the Berlin and Kay (1969) colour categories are universal and that differences in colour names between cultures are unimportant to the underlying cognitive structure; this was denied by subsequent evidence that colour categories were closely linked to the colour terms in the speaker’s language (Davidoff et al., 1999; Kay & Regier, 2006; Roberson et al., 2000; Winawer et al., 2007). However, many of the early experiments on CP (Bornstein & Korda, 1984; Roberson et al., 2000) asked for an identity match in memory. It could be that observers were merely matching the names of colours and that would be sufficient to account for what appeared to be colour CP (Brown, Lindsay & Guckes, 2011; Munnich & Landau, 2003). One clear denial that naming by itself can provide an explanation for CP comes from a patient who could not name colours yet still showed CP (Roberson et al., 1999). Furthermore, recent research does not involve memory and/or avoids overt naming. For example, Winawer et al. (2007) asked for an identity match with the target and samples present, and Daoutis, Pilling, and Davies (2006) introduced a visual search paradigm with rapid presentations, now commonly in use, that asks the observer simply to find the odd-one-out colour.

It has also been argued that CP is merely apparent because it really derives from the metric
used to produce colour stimuli. It is most usual for psychologists to assess equal separation from the Munsell system (Munsell, 1905; Newhall, Nickerson, & Judd, 1943), which is a metric where colours are intended to be equally spaced based on human perceptual similarity judgements. The generality of the claim for equal separation is found justified from threshold discrimination studies (Davidoff & Fagot, 2010; Roberson et al., 2009). But, as has been pointed out by Brown et al. (2011), the assumption about equal spacing in the Munsell system falls down particularly for colours that cross the blue-green boundary. However, a defect of spacing within the Munsell system should predict that speakers of languages that use the same word for green and blue should nevertheless show CP; they do not (Roberson et al., 2000). Moreover, defects of spacing within the Munsell system for blue-green would not explain why CP is found for other category boundaries (Goldstein, Davidoff, & Roberson, 2009; Roberson, Pak, & Hanley, 2008), or for CP distinctions between blues where languages make that distinction (Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009; Winawer et al., 2007) or for newly acquired categories within our normal green category (Ozgen & Davies, 2002). For the present experiments, the choice of metric is somewhat irrelevant as our main point will be to show that the same metric can produce both perceptual and categorical judgements of colour similarity.

After the acquisition of colour terms, not only are judgements made by categorical similarity but they are done so automatically (Clifford, Holmes, Davies, & Franklin, 2010; Fonteneau & Davidoff, 2007; Holmes, Franklin, Clifford, & Davies, 2009; Siok et al., 2009; Tan et al., 2008; Thierry et al., 2009). For example, Thierry et al. (2009) showed that the two Greek colour terms distinguishing light and dark blue leads to greater and faster perceptual discrimination of these colours in native speakers of Greek than in native speakers of English. The visual mismatch negativity, an index of automatic and preattentive change detection, was similar for blue and green deviant stimuli during a colour oddball detection task in English participants, but it was significantly larger for blue than green deviant stimuli in native speakers of Greek. The categorical response was observed early in the ERP trace (i.e., P1 normally associated with perceptual differences) as has also been shown by Holmes et al. (2009).

All of the tasks commonly in use in CP research might have been thought to emphasise judgements of perceptual similarity but they instead promote judgements of categorical similarity. One might also note that they rely on implicit judgements to assess colour similarity. The connection of categorical similarity to implicit processing is also drawn from a clinical case study where an aphasic patient still made categorical errors in an identity match despite being unable to arrive explicitly at colour categories (Roberson et al., 1999). A contrary example is found in Webster and Kay (2011), where very little CP was found in the implicit task of orientation judgement from colour segmentation; nevertheless, there is a reasonable case to be made that the default position in humans is to use, when available, categorical rather than perceptual similarity in tasks that ask for colour similarity judgements.

In the field of concept development there is much discussion on the relative roles of perceptual and categorical similarity (see Sloutsky, 2010). For colour categories, there is much less discussion. However, the matter deserves some attention and the present paper considers tasks that promote a preference for making perceptual rather than categorical judgements. Thus, the main aim of the present paper was to find conditions under which categorical similarity would not be used in judgements of colour similarity. Explicit tasks are more likely to be susceptible to strategic variation and it is these we examine first. For example, different groups of observers were asked to sort a large number of colours that covered most of colour space (Roberson, Davies, Corbett, & Vandervyver, 2005). The colours included good examples of Western colour categories and the outcome for Western observers could be explained by sorting by names (categorical similarity). However, speakers of many African (including Himba—see later) and Australasian (including Berinmo; Davidoff et al., 1999) languages with fewer colour names did not take this opportunity; they made many more categories presumably based on perceptual similarity. For our study, we will use
the explicit task of matching-to-sample (MTS) as there is a surface similarity to the implicit tasks that produce CP. In colour MTS (Fagot, Goldstein, Davidoff, & Pickering, 2006), the observer is given two sample colours drawn from different colour categories (e.g., green or blue) and asked to assign intermediary (target) colours to one of the samples. In implicit versions of that task, the use of categorical similarity appears to be mandatory but will it be so in MTS?

In Experiment 1, we alter the range of stimuli that cross the green–blue boundary. We ask whether the importance of an existing colour category boundary might be so great that the division of stimuli would still be categorical; that is made at the boundary irrespective of where that boundary was placed in the range of stimuli (see Wright & Cumming, 1971, for an argument and data supporting that outcome with respect to colour matching in the pigeon). There are several reasons to expect that result. First, we have seen that categorical similarity appears to be the default similarity mechanism. Second, categorical matching might be favoured because our choice of colours in the Munsell system might give a discontinuity at the boundary (Brown et al., 2011), and lastly the observer might implicitly or explicitly name the colours. Alternatively, we ask whether the division of stimuli could be determined by perceptual similarity and hence not necessarily made at the colour boundary. In particular, it could be made at the middle of the range as Komarova et al. (2007) point to “the existence of a universal human tendency to organise stimulus domains using polar opposition” (p. 363); this line of argument based on the tendency to make a binary division is one that we develop here by altering the range of colours presented to the observer.

Range effects are found for many judgements (Poulton, 1989). For example, judgements of size (small vs. large) are adjusted to the endpoints of the range (Parducci, 1965) and there is the well-known effect of central tendency where judgements of, say, size tend towards some estimate of the middle of the range (Helson, 1947; Hollingworth, 1910). We ask these questions concerning range effects for two very different populations: Westerners and a population (Himba) from a remote savannah region of Namibia. The Himba have only five colour names (Roberson et al., 2004) and make many more categories than Westerners in colour sorting tasks (Roberson, Davies, et al., 2005). With these two disparate populations, we could determine if the use of perceptual similarity (binary division) was widespread in the explicit task of MTS despite their many differences not least being a quite different colour vocabulary.

In Experiment 2, we ask about the role of focal colours. Colour categories have been argued to arise from representations based on focal colours (Rosch, 1975) where colours are drawn towards the prototype to which they belong—the magnet effect of prototypes (Kuhl, 1991). However, colour categories have also been argued to arise from demarcation at their boundaries (Ashby & Gott, 1988; Goldstone, 1994) and it is the norm to find enhanced discrimination at the colour boundary rather than close to the focus (Bornstein & Korda, 1984; Gilbert et al., 2006; Roberson et al., 2000; Winawer et al., 2007; though see Roberson et al., 2009). In Experiment 2, we included a focal colour in the range of colours in a similar MTS task to that of Experiment 1. In previous research using a face MTS task, Angeli, Davidoff, and Valentine (2008) found that a distinctive sample face produced a significant deviation from the equal division of the range with the psychological midpoint of the range now shifted away from the prototype (typical face). Here we ask whether a focal colour stimulus positioned away from the midpoint of the range would also affect the division of stimuli.

In Experiment 3, we turn our investigation of range effects to an implicit task typically used to assess categorical similarity for colour. In Experiment 3, we assessed the availability of colour categories in a visual search task where the range of colours allowed the existing colour boundary to fall at the midpoint of the range or was shifted from it to the same extent as in Experiment 1.

**EXPERIMENT 1: COLOUR MTS WITH VARYING COLOUR RANGES**

Colours provide some uncertainty of their allocation at category boundaries (Rosch, 1975) but the boundary used in Fagot et al. (2006)—that between blue and green—is agreed to be reliable and fairly sharp (Bornstein & Korda, 1984; Boynton & Gordon, 1965; Roberson et al., 1999). In the present study, the importance of the boundary colour in a MTS paradigm will be assessed by changing the range of colours in the matching task. The change (shift) would make the range more blue by moving the dominant
wavelength of the two samples in the shortwave direction, or more green, if moved in the longwave direction. In both cases, and very important, the midpoint of the new range would not normally be acceptable as the boundary colour between green and blue. Thus, the question posed is whether we would still maintain an equal division of the colours, suggesting judgements of perceptual similarity, or whether the observer would make judgements based on categorical similarity and thereby produce an unequal division based on the existing colour boundary (Midpoint Westerners: MW) between green and blue.

To examine the view that there might be a universal tendency to make binary divisions in perceptual continua in MTS, we also examined the effects of range in a remote population. The Himba of Namibia are an extremely remote seminomadic population of animal herders estimated between 20,000/50,000 in what has been described as the last wilderness in southern Africa (Namibian Government Statistics, 2004). They do not have words that map directly on to our terms in their five colour term language but they nevertheless have sharp boundaries between their terms (Roberson et al., 2004). Given that the Himba do not have a boundary between green and blue, it might be thought unsurprising if range effects were found for those colours. More interesting would be if range changes were found to affect one of their own established colour boundaries. A comparable MTS experiment was therefore undertaken on a colour range that crosses the boundary between two of the Himba colour terms. Roberson et al. (2004) demonstrated a sharp boundary for colours that Westerners would call “green” on both sides of their boundary. To the yellow side of that boundary colour, the Himba use the term “dumbu” and to the blue side “buri”. The comparable experiment used the range with their boundary colour in the middle (Midpoint Himba: MH) and also used shift conditions where the range of colours is moved either towards the shortwave or longwave end of the spectrum. The design allowed us to see if the Himba divided the range at the midpoint and then investigate the effect of the shifts to see whether the allocation of colours to the sample was made on the same basis as for Western observers. For both Western and Himba observers we used the modelling procedures in Fagot et al. (2006) to establish where in the range observers placed their boundary.

One further question we address is whether the names given to the colours are affected by the range. Naming is of course an explicit task but it differs from MTS in that the categorical aspect of colour similarity ought to be built into the response. Indeed, the evidence to date suggests that naming is unaltered after changing the range of colours (Mitterer & de Ruiter, 2008).

**Method**

**Participants.** Different Western observers were used for three colour ranges (MW [no-shift], green-shift, and blue-shift). The eight observers (and data) for the MW (no-shift) range were taken from Fagot et al. (2006). Four different French (three women and one man; age 24–29 years, mean 26.7 years) observers and eight English (four women and four men; age 24–34 years, mean 29.7 years) observers saw the colours of the continuum shifted to make them more green (green-shift). A further different 12 observers from the same places and with the same gender distribution (French: age 23–29 years, mean 26.5 years; English: age 24–32 years, mean 28.3 years) saw the continuum shifted to make the colours more blue (blue-shift). The French observers were studying in Luminy, Marseille, France and the English were students at Goldsmiths, University of London, UK. All were screened for red–green colour vision abnormalities with the Ishihara colour vision test (Ishihara, 2001) and paid for taking part in the experiment.

For Himba observers, there were three groups of 12 participants. The first group who conducted the study with their boundary in the middle of the range (MH) contained four men and eight women, varying in age from approximately 18 to 40 years, mean 30 years. The second group (nine women and three men; age range approximately 20–60 years, mean 39 years) saw colours shifted towards yellow (dumbu-shift). The third group (10 women and two men; age range approximately 17–45 years, mean 27.6 years) saw colours shifted towards blue (buri-shift). All observers were screened for colour vision abnormalities along the red–green confusion axis with the City Colour Vision Test (Fletcher, 1980)—a test that does not require number identification—and paid with small quantities of maize flour.

**Stimuli.** All experiments used Munsell (1905) colours at Value (Brightness) 5 and Chroma...
(Saturation) 6 with 2.5 Munsell hue units between each adjacent colour. The MW group of Western observers in Fagot et al. (2006) saw 12 colours going from 2.5G to 10B centred around the known boundary for green and blue (between 5BG and 7.5BG; Roberson et al., 1999). For the shift condition, 11 stimuli were used to make only one stimulus in the median position. The 11 stimuli used for the green-shift continuum for Western observers were 10GY to 5B; i.e., the range shifted 5 Munsell steps in a greenwards direction. The 11 stimuli for the blue-shift continuum were 7.5G to 2.5PB; i.e., the range shifted 5 Munsell steps in a bluewards direction. The colours 10G, 2.5BG, 5BG, 7.5BG, 10BG, and 2.5B were common to all groups. Colours for the Western observers were generated on a computer screen using colour notation software (Roberson & Davidoff, 2000) and were calibrated with a CS100 colour gun so as to assure correct colour appearance. Colour stimuli subtended 6.4° by 6.4° visual angle.

The Himba boundary between dumbu and buru is between 7.5GY and 10GY (Roberson et al., 2004). For the Himba, the MH group saw 12 colours going from 2.5G to 10B centred around the known boundary for green and blue (between 5BG and 7.5BG; Roberson et al., 1999). For the shift condition, 11 stimuli were used to make only one stimulus in the median position. The 11 stimuli used for the green-shift continuum for Western observers were 10GY to 5B; i.e., the range shifted 5 Munsell steps in a greenwards direction. The 11 stimuli for the blue-shift continuum were 7.5G to 2.5PB; i.e., the range shifted 5 Munsell steps in a bluewards direction. The colours 10G, 2.5BG, 5BG, 7.5BG, 10BG, and 2.5B were common to all groups. Colours for the Himba observers were standardised. The decision to use Illuminant C under which the Munsell colours were placed into a solar-powered lightbox was not based on a concern that viewing distance remained approximately equal to 49 cm. Control and randomisation of conditions were achieved through purpose-made programs written in E-Prime V 1.0 (Psychology Software Tools, Inc.). The Himba observers were tested manually under the shade of a tree and colours were placed into a solar-powered lightbox giving Illuminant C under which the Munsell samples are standardised. The decision to use the lightbox was not based on a concern that computer screens could be too unfamiliar to the Himba. They are also unfamiliar with printed material but we have found, in examining their shape perception, identical outcomes for material when presented in that form or on a computer screen (Davidoff, Fonteneau, & Goldstein, 2008; De Fockert, Davidoff, Fagot, Parron, & Goldstein, 2007). Our reason for using the lightbox was that colours are exactly reproduced; we could not ensure that a computer screen maintained colour calibration in the field. The two procedures give identical outcomes for Western observers (Roberson & Davidoff, 2000; Roberson et al., 2000).

Procedure. To initiate the testing, Western participants pressed the spacebar. A square-shaped sample stimulus then appeared with 4.5° of lateral eccentricity on the right or the left of the screen followed 500 ms later by a cursor and two patches of colour for comparison with the sample. In balanced order, one colour square appeared on the top, and the other on the bottom half of the screen. Observers indicated which target was more similar to the sample with a joystick. There was no time limit for responding. Response choices and latencies were recorded.

For the MW group, the training procedure used only the 2.5G and 10B stimuli as sample and target stimuli; the green-shift group used only 10GY and 5B and the blue-shift group 7.5G and 2.5PB. Training sessions for all groups was as in Fagot et al. (2006) for the MW group. Sessions were composed of 96 randomly ordered identity matching trials, resulting from a completely balanced stimulus identity by stimulus position design. Observers were required to reach a criterion level of 80% correct in training trials before being tested; they were reinforced at this stage receiving yes (oui)/no (non) feedback.

Experimental test sessions were composed of identity and similarity MTS trials. In the identity trials, the target was identical to one of the sample stimuli. For the no-shift (MW) group (see Fagot et al., 2006), in the similarity matching trials, the target was randomly chosen from the 10 intermediary colours, going from 5G to 7.5B. For all trials, the colour-comparison stimuli were 2.5G and 10B. All possible combinations of stimulus position were given equally often in a random order. The two stimuli used on identity matching trials were each shown 15 times, and the 10 stimuli used on similarity matching trials were each shown once, for each combination of sample and comparison stimulus position, resulting in a total of 160 trials per block. The proportion of identity and similarity matching trials (3:1) was chosen so as to maintain attention during the test.
Ten test blocks were required for each observer who was allowed breaks on every occasion after completion of the first three test sessions. A similar procedure was adopted for the green-shift and blue-shift groups except that the target was randomly chosen from the nine intermediate colours, going from 2.5G to 2.5B (green-shift) and 10G to 10B (blue-shift).

A similar but manually controlled procedure was carried out for Himba observers. The endpoints of the chosen range (MH [no-shift], dumbu-shift, or buru-shift) were used as training stimuli. They were shown individually, one at a time in random order and observers were asked in Himba to look at the colour carefully. For training, the colour was then removed and, with an interval of a couple of seconds, followed by the administration of the choice comprised of the two endpoints (e.g., 5Y and 2.5BG). The observer was then asked, by an interpreter, which colour he/she had seen. The observer was trained until he/she got all the practice trials right four times consecutively.

For the MH (no-shift) range, the test phase of the MTS experiment used the 10 Munsell colours (7.5Y to 10BG) intermediate between the endpoints of the range. A shortened procedure was carried out with the Himba in order to maximise data collection from observers who might not be able to return for testing. The test phase in the MH and shift MTS tasks each consisted of 200 trials with every colour presented 20 times as a target. The sample choices were now present at the same time as the targets and placed on a card that could be reversed to keep the colour of the top and bottom sample randomised. A similar procedure was adopted for the test trials in the shift conditions. The instructions to the Himba were to match the test stimulus to the target that was more similar; this instruction induced no strategic bias, i.e., it did not emphasise categorical or perceptual similarity judgements.

After the MTS, both Himba and Western observers were asked to name the colours presented during the experiment. Western observers were asked to name the colours individually shown on the screen in random order; this was repeated five times. Himba observers named the colours once in random order in the lightbox.

**Results of MTS**

*Training*

For Western observers, one session sufficed for them to learn the MTS rule. The Himba observers needed at most 16 identity matching trials to reach criterion for the test.

*Modelling the data*

Following the same rationale as Fagot et al. (2006), we modelled the data to determine the most likely strategy used by the participants in the MTS task. We fitted the two-parameter (human) decision bound (DB) model from Fagot et al. to the similarity matching data for each observer. The DB model proposes that the observer sets a linear boundary at a particular value \( (k) \) within the range of the perceptual response values created by the probe stimuli. For each probe trial, the matching response to training stimuli (e.g., 2.5G [Green] or 10B [Blue] for Western observers) is determined according to whether the perceptual response value of the probe falls above or below \( k \). In the no-shift conditions, 10 probe stimuli were used and thus the value of the boundary parameter can be represented by a single value in the range 1–10; the nine probe stimuli for the green-shift (dumbu-shift) and blue-shift (buru-shift) conditions are represented by the range 0–8 and 3–11, respectively. This boundary is employed for all probe stimuli but its application is susceptible to noise (\( \sigma^2 \)). The noise parameter jointly accounts for both the error associated with the placement (or application) of the decision boundary and also the perceptual representation of the probe values, across trials. The two-parameter (\( k \) and \( \sigma^2 \)) DB model was fit, for each observer individually, to the probability of making a particular matching response (i.e., “green” or “dumbu” for Western and Himba observers, respectively) for each of the probe stimulus types (nine or 10 for the shift and no-shift conditions, respectively). Maximum likelihood estimates of the two parameters were found using a constrained iterative routine in MATLAB (see Fagot et al., 2006). The fit of the two-parameter model was compared with a perfect saturated model using Likelihood Ratio (LR) test statistics (for a review of this procedure see Maddox & Ashby, 1993). This saturated model used nine or 10 parameters (the observed probabilities of “green” or “dumbu” responses for
each of the probe stimulus types) for the shift and no-shift conditions, respectively, and was (trivially) able to predict the observed probabilities perfectly. When the LR test statistic is nonsignificant (i.e., below the 95th percentile of the \( \chi^2 \) distribution, with degrees of freedom equal to the difference in the numbers of parameters between the two models, i.e., \( df = 7 \) or \( df = 8 \) as appropriate), then the simpler two-parameter model is not significantly worse than the perfect model, and it can be deemed an extremely good fit.

The two-parameter DB model was fitted to each individual’s data and yielded an excellent fit in most cases. For 14 of the 20 Western and Himba observers in the no-shift condition, the two-parameter model did not yield a significantly poorer fit than the saturated model (LR test values ranged from 0.92 to 13.67; 7 \( df \); .057 < \( p \) < .996). For these six observers, the deterioration of fit from the saturated model was moderate (LR test values ranged from 14.73 to 19.06; 7 \( df \); .008 < \( p \) < .040). There was no significant difference in mean value of the noise parameter across the shift conditions for either Western observers, \( F(2, 29) = 1.96 \), \( p > .15 \), \( \eta^2_p = .12 \), or for the Himba, \( F(2, 33) < 1 \), \( \eta^2_p = .05 \).

**Range effects**

We examined the effect of shifting the range of the probe stimuli on the modelled decision boundary (\( k \)) parameter separately for the green–blue and dumbu–buru ranges.

**Green–blue continuum.** For Western observers, the mean value of the boundary estimate (\( k \)) was entered into a three level (group: no-shift [MW] vs. green-shift vs. blue-shift) ANOVA and revealed a significant main effect of group, \( F(2, 29) = 51.28 \), \( p < .001 \), \( \eta^2_p = .78 \); see Figure 1. Planned contrasts showed that, relative to the no-shift group, the mean estimated boundary values for the green-shift (3.01, \( SD = 0.77 \)) and blue-shift
groups were significantly shifted, respectively, greenwards, \( t(29) = 5.53, p < .001, d = 2.03 \), or bluewards, \( t(29) = 3.47, p < .004, d = 1.29 \); both contrasts Bonferroni corrected for two comparisons.

In Figure 1 it can be seen that the boundaries for all three groups were placed near the middle of the ranges. For the no-shift condition, the middle of the range coincided with the existing colour boundary between green and blue. Thus, perhaps not surprisingly, the observed value of \( k \) did not differ from the predicted value of 5.5. The mean value of the \( k \) parameter for the Western (\( M_W \)) no-shift group was 5.15 (4.49–6.05, \( SD = 0.56 \)) and gave \( t(7) = 1.76, p > .05, d = 1.33 \). The midpoints between the two training stimuli used in the two shift conditions represented a shift of \(-1.5\) (green-shift) and \(+1.5\) (blue-shift) units (along the \( k \) scale) from the midpoint of the no-shift condition. Based upon the observed midpoint of the no-shift group (5.15) one would expect a mean \( k \) parameter values of 3.65 and 6.65 for the green-shift and blue-shift conditions, respectively, if decisions were based on the middle of the range. The mean of the blue-shift group (6.50) was not significantly below 6.65, \( t(11) < 1 \). The mean of the green-shift group (3.01) was, in fact, lower than 3.65, \( t(11) = 2.85, p < .05, d = 1.72 \) in a direction away from the existing colour boundary (see Figure 1).

To examine whether the shift conditions resulted in a corresponding latency difference, the six colours common to the three continua used for Westerners were submitted to analysis. Latencies of the green-shift and blue-shift data were prorated to average performance levels of the no-shift (\( M_W \)) group and are shown in Figure 2. A 3 (group: no-shift [\( M_W \)] vs. green-shift vs. blue-shift) \( \times 6 \) (common probe colour: 10G, 2.5BG, 5BG, 7.5BG, 10BG, and 2.5B) ANOVA was carried out on log latencies. The analysis revealed group differences, \( F(2, 29) = 7.32, p < .005, \eta_p^2 = .336 \) in the context of a main effect of common probe colour, \( F(5, 145) = 4.86, p < .005, \eta_p^2 = .144 \), and an interaction, \( F(10, 145) = 20.63, p < .0001, \eta_p^2 = .587 \). It is clear from Figure 2 that the three groups have their peak latencies at different colours (i.e., the middle of each range). When the continuum was shifted greenwards, peak latencies were found to correspond to a shift in that direction (\( p < .0001 \)); when the continuum was shifted bluewards, peak latencies were then found to have shifted in that direction (\( p < .002 \)).
**Dumbu–buru continuum.** For Himba observers, the mean value of the boundary estimate \((k)\) was entered into a three-level (group: no-shift \([M_H]\) vs. dumbu-shift vs. buru-shift) ANOVA and again revealed a significant main effect of group (see Figure 3) on the estimated decision boundary, \(F(2, 33) = 13.04, p < .001, \eta^2_p = .44\). Relative to the no-shift group, the mean estimated boundary value for the dumbu-shift group (4.13, \(SD = 0.83\)) was significantly shifted towards the dumbu end of the colour continuum, \(t(33) = 2.77, p < .02, d = 0.96\). The degree of shift in the boundary estimate towards the buru end of the colour continuum in the buru-shift group (5.99, \(SD = 1.16\)) was also in the predicted direction, \(t(33) = 2.33, p < .06, d = 0.81\), both contrasts Bonferroni corrected for two comparisons. Again, not surprisingly, the mean value of the \(k\) parameter for the Himba \((M_H)\) no-shift group (\(M = 5.14, range 4.03–6.03, SD = 0.59\)) did not differ from the middle of the range, \(t(11) = 2.12, p > .05, d = 1.28\). More critical, based upon this observed midpoint of the no-shift group, the mean values of the \(k\) parameter for the dumbu-shift group (4.13) and buru-shift group (5.99) were not significantly different from their respective range midpoints of 3.64 and 6.64, \(t(11) = 2.03, p > .05, d = 1.22\), and \(t(11) = 1.96, p > .05, d = 1.81\), respectively.

**Discussion of MTS**

On each MTS trial, the structure of task was to present three stimuli of which two were from the same colour category just as it is in tasks that produce CP. Yet the MTS produces judgements made on perceptual rather than categorical similarity. The two sets of human observers could hardly be more different in terms of their literacy and culture, but in a striking fashion they both showed judgements based on perceptual similarity (binary division) in the MTS. Both Westerners and Himba made relatively sharp equal divisions in all ranges irrespective of the position of their colour boundary. Observers did not know at the outset of the test the exemplars for which judgments were required, yet, with relatively few trials and with only one minor exception, both groups apportioned the exemplars equally to the two sample stimuli. Furthermore, the resulting middle of the range boundary overrode a category boundary that in other circumstances has been found to be robust and well-established.
Admittedly, we tested the phenomenon with a restricted range of colours, but, even so, the effects were strong. The critical difference in the MTS procedure could be that the observer is presented with two constant target colours and a third (sample) colour that takes many positions in between. Some of those sample colours were very similar to one of the targets and this may be the critical factor that encourages the use of perceptual similarity.

It might seem obvious that, in the absence of other biases, observers would make equal divisions of the continuum, but it is worth commenting upon. For example, it might explain why colour categories are of roughly equal size (see also the similar prediction in Komarova et al., 2007) and have the extra benefit of being the most efficient procedure for adding new colour terms to a speaker’s language (Jameson, 2005). In a more general context, the tendency to make divisions in perceptual continua could also be the basis of dichotomies seen in so much of human cognition. Indeed, for some social theorists, dichotomies are at the core of human thought (e.g., Lévi-Strauss, 1969) and even been argued to be hard-wired into the brain (see Rose, 2005).

**Results of naming.** We examined the naming data to see if range changes also affected the names given to the colours. For Westerners, a naming boundary was found for each observer by determining the first point along the green/blue colour continuum at which the number of green responses dropped below chance levels (50%). The same general procedure was applied to the Himba but they were presented with only one example of each colour used in the respective MTS tasks and were not restricted in their possible colour naming response. To aid comparison across shift conditions, we represented the colour naming boundaries according to the $k$ parameter used for the modelling. Thus, the midpoints of both the Western and Himba colour probe ranges, i.e., the established colour boundaries (green–blue: 5BG–7.5BG, and dumbu–buru: 7.5GY–10GY, respectively), are represented by a value of 5.5$k$ across all shift conditions. In Table 1, we give the naming boundaries for the Westerners for the three ranges used in green–blue and, for the Himba, the three ranges used for dumbu–buru. It is clear that the range made only the most marginal difference to the names given to the colours. For Westerners, there was no hint of any significant difference between the colour naming boundaries across the three groups, $F(2, 29) < 1$, $\eta_p^2 = .057$. The mean boundary value of each group was close to a $k$-value of 5.5 (see Table 1) and all lay within the usual boundary range of green–blue; i.e., between 5 BG (5$k$) and 7.5BG (6$k$).

Colour naming is more variable in the Himba (Roberson et al., 2004) with some observers preferring to use cattle terms for the colours. On that basis, eight observers were excluded from analysis (one from the no-shift, four from the dumbu-shift, and three from the buru-shift groups). The resulting analysis gave no significant difference between the colour naming boundaries across the three groups, $F(2, 25) < 1$, $\eta_p^2 = .060$. The mean boundary values of the dumbu-shift and buru-shift groups were both close to the midpoint 5.5. While not significantly different from the two shift conditions, the mean colour

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>No-shift (MW)</th>
<th>Green-shift</th>
<th>Blue-shift</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>k-value</strong></td>
<td><strong>Munsell colour</strong></td>
<td><strong>k-value</strong></td>
</tr>
<tr>
<td>Westerners green vs. blue</td>
<td>5.75 (1.39)</td>
<td>6.88BG</td>
</tr>
<tr>
<td>Himba dumbu vs. buru</td>
<td>4.86 (0.92)</td>
<td>4.62BG</td>
</tr>
</tbody>
</table>

For all ranges, a value for $k$ of 5.5 represents the middle of the range. For Westerners, colour naming was obtained from the no-shift (MW), green-shift, and blue-shift continua; for the Himba it was obtained from the no-shift (MH), dumbu-shift, and buru-shift continua. No boundaries are reliably different from $k = 5.5$. 

(Bornstein & Korda, 1984; Roberson et al., 1999).
naming boundary for the no-shift group was somewhat below 5.5.

Discussion

In summary, when asked to name the colours, the observers assign the colours in a way consistent with categorical similarity. In MTS, observers assign colours consistent with perceptual similarity. It is perhaps remarkable given the claim by some authors of automatic colour naming as an explanation of colour CP (Munnich & Landau, 2003) that observers do not turn to naming in the MTS but clearly they do not. In the next two experiments we explore the generality of our results. In Experiment 2, we examined the effects of the inclusion of a focal colour to rule out a potential artefact in the MTS paradigm.

EXPERIMENT 2: COLOUR MTS WITH A RANGE INCLUDING THE FOCAL COLOUR

In Experiment 1, we assumed that the observer was obeying the instruction to assess the similarity of the sample colour to the targets. However, the results of Experiment 1 could reflect neither a perceptual nor a cognitive judgement but merely a response bias to make equal responses to each sample in MTS. We therefore wish to rule out this alternative in Experiment 2. Evidence from MTS tasks with other domains (Kuhl, 1991) suggests that the division of stimuli might not always be equal where there is the inclusion of focal examples. Indeed, one might have explained the one case of an unequal division in Experiment 1 in this way as the shift condition now included a colour close to focal green.

Focal colours have a psychological importance. They are earlier and more reliably established than boundary colours (Mervis, Catlin, & Rosch, 1975), better recognised in short-term memory (Rosch Heider, 1972), more effective in paired-associate learning (Rosch Heider, 1972), and are more efficiently activated by the colour word (Rosch, 1975). It is uncertain whether Western focal colours have a universal importance as argued by Kay and Regier (2006) or whether each language forms its own categories around its own best examples (foci) of colours given the same name (Roberson et al., 2000; Roberson, Davies et al., 2005). However, irrespective of the origin of focal colours, it seems important to ask whether the equal divisions of the colour ranges found in Experiment 1 would be maintained if a focal colour were included in the similarity judgements.

In Experiment 2, we therefore repeated the MTS experiment with a range of colours that would all normally be called green but included focal green (Berlin & Kay, 1969) positioned towards one end of the range. A dominant view (Kuhl, 1991) is that a typical stimulus (prototype) as an endpoint will act like a magnet and draw stimuli towards it thus causing the boundary to be drawn further from the prototype. An alternative view is that atypical stimuli have larger attractive fields with the opposite prediction about boundary position (Tanaka, Giles, Kremen, & Simon, 1998).

If judgements of similarity were more likely to be made to the focal (typical) sample (Angeli et al., 2008; Kuhl, 1991) or to the atypical sample (Tanaka et al., 1998), then clearly the midpoint of the MTS judgement would be altered by the inclusion of the focal colour. An asymmetric response would also show that the results of Experiment 1 reflected a perceptual judgement and were not merely a response bias to make equal responses to each sample.

Method

Participants. Twelve observers (seven women and five men; age 18–35 years, mean 23.30 years) were recruited at Goldsmiths, University of London, UK. All had normal colour vision for the red-green confusion axis (Ishihara, 2001) and were paid for participating in the study.

Stimuli. The stimuli were derived from Munsell colours and calibrated as in Experiment 1 but with Value (Brightness) 2.5 and Chroma (Saturation) 6 and a linear spacing of 1.25 Munsell hue units between each adjacent colour. Darker stimuli were chosen to include an example of focal green and smaller Munsell separation to ensure all colours would be called green. Eleven stimuli were generated ranging from 7.5GY to 10G and subtending 6.4° by 6.4° visual angle when presented on a computer screen. The aim was to use a range of colours including a focal colour towards one end of the range. Therefore, we
included focal green, which is at approximately 10GY (Davidoff et al., 1999) in our range of greens. All colours in Experiment 2 would be normally called “green” (Roberson et al., 1999).

Procedure. The procedure was identical to Experiment 1 except for the use of a different colour range and, instead of using a joystick, observers simply pressed one of two buttons arranged vertically to correspond with a cursor movement of “up” or “down”. The training procedure used only the two endpoint colours of the range 7.5GY and 10G as the sample and target stimuli. For the similarity trials, the target was one of the nine intermediary colours from 8.75GY to 8.75G. For all trials, the sample stimuli were 7.5GY and 10G. The task was again to assess which of the two sample colours was more similar to the target.

Results

Again, all participants had learnt the MTS rule by one session. We examined the effect of the range of the probe stimuli on the modelled decision boundary (k) parameter. The average value of k was found to be at 2.19 G ($SE = 0.16$). A one-sample $t$-test indicated that the boundary was statistically significantly different from the midpoint of the range (3.75G), $t(11) = 9.57, p < .001$ (see Figure 4), shifted greenwards (i.e., towards the end of the continuum containing focal green at 10GY). Using log transformed latency data, a one-factor ANOVA on target colour (11 levels) indicated that there was a significant main effect of target colour, $F(10, 121) = 7.97, p < .001$. Planned simple contrasts revealed that latencies for 1.25G were not significantly longer than for 2.5G ($p > .1$) but they were significantly longer than for all the other colours (all $p s < .01$; see Figure 4). These latencies reflect the choice data where the average boundary was at 2.19G (i.e., between 1.25G and 2.5G).

Discussion

In Experiment 2, we found that the categorical decision did not give a boundary at the middle of the range. The asymmetric response indicates that the results of Experiment 1 were not merely a response bias to make equal responses to each sample. The boundary in the MTS was drawn closer to the focal colour and this would seem counter to the argument in Kuhl (1991) that similarity judgements were more likely to be made to the typical (here focal) sample. It is worthwhile noting the similar outcome of the inclusion of a focal colour when observers are asked for an explicit colour (naming) categorisation (Hampton, Estes, & Simmons, 2005). In their
study, Hampton et al. (2005) found that observers would more likely categorise a borderline blue-purple colour as purple if it were presented alongside a typical blue. In effect, the category boundary in their study was shifted towards the focus as here in Experiment 2. Hampton et al. interpret their data as showing that contrast is an important determinant of (colour) categorisation in simultaneous comparisons (see also Stewart & Brown, 2005). In their categorisation tasks, the observers in Hampton et al. had to decide whether a colour matched a name. Their observers’ borderline judgements were considerably affected by the presence of a focal colour by making it less likely that borderline colours were given the same name as the focal colour.

It is not our primary concern to investigate the role of typical stimuli in the allocation of similarity judgements. Our primary concern is to examine tasks where the range may or may not override existing colour boundaries and hence produce judgements based on perceptual similarity. However, one implication that could be drawn from Hampton et al. (2005) is that even colour similarity assessed by naming might depend on the range of stimuli. One of the aims of Experiment 3 was to examine whether that is the case. In Experiment 3, we considered a colour CP task; such tasks have been claimed to show an automatic activation of colour categories (Fonteneau & Davidoff, 2007) and so might be immune from range effects.3

EXPERIMENT 3: VISUAL SEARCH WITH VARYING COLOUR RANGES

We used a visual search paradigm successfully employed in several recent studies of colour CP (Daoutis, Franklin, Riddett, Clifford, & Davies, 2006; Daoutis, Pilling, & Davies, 2006; Drivonikou et al., 2007; Franklin et al., 2008; Gilbert et al., 2006). These studies have found that search times were quicker for an odd-one-out colour embedded within colours of a different category (cross-category targets) than embedded within colours from the same category (within-category targets). If these effects were in some part due to a division of the range of colours at the midpoint, then some sort of categorical division—perhaps even CP—should be obtained at the midpoint of any range of colours. Experiment 3 contrasted a range of colours with the midpoint at the green-blue boundary with another range of colours where the midpoint was green-shifted to the same extent as in Experiment 1. For the green-shift range, we compared a group who had prior exposure of the range to another that had no prior exposure. We included the former group in order to maximise the green-shift participants’ exposure to the green-shifted stimulus range, and therefore the possibility of a range effect.

Method

Participants. The observers were tested on the green–blue (no-shift) set or the green-shift set. There were 14 observers (10 women, four men, age 19–38 years, mean 23.86 years) in the green–blue condition, 14 observers (six women, eight men, age 20–30 years, mean 24.07 years) in the green-shift condition with prior exposure to the range of stimuli to be used in the visual search task, and 14 observers (12 women, two men, age 19–35 years, mean 23.93 years) without prior exposure. All were recruited at Goldsmiths, University of London, had normal colour vision for the red–green confusion axis (Ishihara, 2001) and were paid for participating in the study.

Stimuli. There were two sets of stimuli. The first (green–blue set; see top half of Figure 5) contained both blue and green patches centred at the green–blue boundary (6.25BG; Roberson et al., 1999), and therefore used the same midpoint as the M_W range of Experiment 1. The second set (green-shift; see bottom half of Figure 5) was shifted five Munsell steps towards the green end of the continuum, with a midpoint now at 1.25BG. All stimulus presentations contained 12 colour patches arranged as in a clock face but offset so that no colours appeared on the vertical meridian; 11 of the colours were identical and one was an odd-one-out. For both sets there were four types of odd-one-out stimulus but always separated from the identical stimuli by five Munsell steps. For both sets of stimuli, two of the types were drawn from the middle of the range (“middle”;

3 Motivation for Experiment 3 came in part from a talk given by Dr. Oliver Wright at the Progress in Colour Studies 2008 conference held in Glasgow, UK (Wright, 2011). He noted that the midpoint of the range of blue/green colours used in the visual search study of Gilbert et al. (2006) did not correspond precisely to the usual boundary given for that colour range.
see lower pairs in each half of Figure 5) and the other two were drawn from the ends of the range ("end"; see upper pairs in each half of Figure 5) For the green–blue (no-shift) set, the middle types would normally be regarded as cross-category; they used 2.5BG as the odd-one-out in 7.5BG (or vice versa) and likewise for 5BG and 10BG. The end types would normally be regarded as within-category; they used 8.75G as the odd-one-out in 3.75BG (or vice versa) and likewise for 8.75BG and 3.75B. The green-shifted stimulus set were equivalent to the green–blue set but shifted by five Munsell steps. The two middle types were (7.5G–2.5BG) and (10G–5BG) and the two end types were (3.75G–8.75G) and (3.75BG–8.75BG). One may note that the end type colours 3.75BG–8.75BG in the green-shift range could be regarded as cross-category given the standard allocation of the green–blue boundary.

**Apparatus.** This was as for the Western observers in Experiment 1 except that the joystick was replaced by a button box to record responses.

**Procedure.** Each observer completed three tests: the main visual search task and two naming tasks covering (1) just their task range of colours, and (2) the entire range of colours across both the no-shift and green-shift stimuli. The no-shift group first completed the visual search task, followed by the naming of their task range and finally the naming of the entire stimulus range. One set of green-shift observers carried out the tasks in the same order. However, in order to maximise the green-shift participants’ exposure to the green-shifted stimulus range, and therefore the possibility of a range effect, the other set of green-shift observers were told that they would see a range of colours that would cover all the colours that might be used in the visual search task. They were then asked to name their task range which was then followed by the visual search task and finally the naming of the entire stimulus range.

Based on Gilbert et al. (2006, Exp. 1), each trial of the visual search task began with a fixation display for 1000 ms, followed by the stimulus display consisting of 12 squares of colours. Only one of these squares (target) was a different colour from the others (distractors), and observers were asked to identify this odd-one-out and respond to whether it was on the left or right side of the screen by pressing the corresponding button of a response box. In Gilbert et al., there was an interest in the hemispheric locus of colour categories; however, as that was not our concern, we did not present stimuli at a rapid speed to prevent eye movements but the display remained on screen until a response was made. Following
the response, a blank screen was displayed for 250 ms before the next trial.

Each stimulus colour served as both target and distracter. Each target was presented equally often at each of the 12 locations, and the order of location and target stimulus was randomised. Thus, each block was made up of 96 trials and participants completed three blocks. Response times and accuracies were collected. A practice block of 12 trials using purple and orange stimuli ensured observers understood the task.

For naming, the observers responded by pressing buttons labelled “blue” or “green”. Each colour was presented individually in a random order on the computer screen in a block of trials. Five blocks were given for the range used in their visual search task. Observers who saw the green–blue (no-shift) set were presented with 17 colours ranging from 8G to 4B in steps of one Munsell unit; the green-shift participants saw 17 colours ranging from 3G to 9BG. For the entire range of colours, participants were asked to make similar responses to 24 colours at one Munsell step from mid range. Thus, each block was made up of 96 trials and each was entered into a 2 (category: middle vs. end) × 3 (stimulus set: green–blue [no-shift] vs. prior exposure green-shift vs. no prior exposure green-shift) mixed ANOVA with the within-participants factor of category, and the between-participants factor of stimulus set. Figure 6 (top) shows that effects of categorical similarity are only found in the no-shift set. There were significant main effects of category, $F(1, 39) = 24.96$, $p < .001$, and interaction, $F(2, 39) = 9.29$, $p < .001$, but no effect of stimulus set, $F(1, 39) < 1$.

Tukey HSD tests on the difference between middle and end performance indicated (see Figure 6, bottom) that the middle-end comparison for the no-shift was significantly different to both the no exposure green-shift ($p < .005$) and prior exposure green shift ($p < .001$) that did not differ from each other ($p > .7$). Again, Tukey HSD tests of the pair types used in the green-shift conditions revealed no sign whatsoever of an advantage for the 3.75BG–8.75BG colours without ($p > .63$) or with ($p > .62$) prior exposure to the range.

**Accuracy.** The accuracies on both stimulus sets were generally high and were therefore inverse sine transformed prior to analysis. A similar mixed ANOVA to that conducted on latencies again indicated (see Figure 6, bottom) that effects of categorical similarity are only found with the no-shift set. There was a significant main effect of category, $F(1, 39) = 7.31$, $p < .01$, an interaction, $F(2, 39) = 10.44$, $p < .001$, but no effect of stimulus set, $F < 1$.

**Naming**

**Range covering observer’s stimulus set.** We examined where observers placed their green–blue naming boundary in the range of stimuli on which they had been tested in the visual search task. The naming boundary was found for each observer as in Experiment 1. The location of the boundary for each observer with respect to the midpoint of their range was calculated in Munsell units. On average, for the green–blue (no-shift) observers, the naming boundary was at 6.50BG (middle of range = 6.25BG). For the no prior exposure green-shift participants, it was at 1.71BG, and for the prior exposure green-shift at 1.36BG (middle of range = 1.25BG) (see Figure 7). One-sample $t$-tests indicated that in all conditions (no-shift, $p > .6$, no prior exposure green-shift, $p > .15$, and prior exposure green shift, $p > .7$) the boundary was not significantly different from the midpoint of the range.

**Large range.** In this naming task, all participants were tested on a large range of stimuli that included colours close to focal examples of both green and blue; the middle of the range was 2.5BG. The average naming boundary of this large range was at 5.07BG for the green–blue (no-shift) participants, 3.71BG for the no
exposure green-shift participants, and 3.14BG for the prior exposure green-shift participants, (see Figure 8).

One-sample $t$-tests indicated that in the no-shift condition, $t(13) = 4.66, p < .001$, and no prior exposure green-shift condition, $t(13) = 2.97, p < .02$, the boundaries were significantly different from the midpoint of the range; however, the prior exposure green-shift condition escaped significance, $t(13) = 1.80, p < .1$. Tukey tests indicated that the no-shift boundary was not different from the no prior exposure green-shift, $p > .05$, although it was to the prior exposure condition ($p < .05$). The prior and no prior exposure green-shift boundaries did not differ ($p > .6$).

The boundary of the no-shift condition did not differ from the assumed green–blue boundary at 6.25BG, $t(13) = 2.14, p > .05$. However, although not reliably different from the no-shift condition boundary, the no exposure green-shift boundary when separately tested against the assumed green–blue boundary was found to be different to 6.25BG, $t(13) = 6.21, p < .01$, as was the prior exposure green-shift boundary, $t(13) = 8.70, p < .001$.

Figure 6. Top: Mean accuracies ($\pm 1 SE$) in each category condition for the visual search task in the no-shift (green-blue) and two green-shift groups. For the no-shift condition, “middle” would correspond to a cross-category pair and “end” to a within-category pair. Bottom: Mean latencies ($\pm 1 SE$) in each category condition in the visual search task for the no-shift (green-blue) and two green-shift groups. For the no-shift (green-blue) condition, “middle” corresponded to a cross-category pair and “end” to a within-category pair.
Discussion

Experiment 3 showed two important different outcomes to Experiment 1. The first difference was that, with the visual search CP paradigm, we found colour similarity based on categorical similarity. However, CP was found only for the green–blue set. The other set did not show CP whether or not participants had prior exposure to the range. An important difference between the paradigms used in Experiments 1 and 3 is that there is an explicit demand for a similarity match in MTS, whereas in the odd-one-out task the category judgement is implicit. Consideration of these issues concerning when colour category boundaries are immune from range effects will be reserved for the General Discussion. The second and surprising difference is that there are now ranges of stimuli for which naming does not conform to what might be expected from existing colour boundaries.

In Experiment 1, naming could be seen as a reliable measure of categorical similarity by being
immune from range effects; in varying degrees this was not the case in Experiment 3. Whereas for the no-shift condition, there was little difference to the names given to the colours, prior exposure to the shift range produced a reliable change in naming for the large range. Also, and more dramatically, the naming changed with a restricted range. When both green-shift groups named the more restricted colour range used in their visual search task, their naming boundary was now clearly shifted to the middle of the range.

The range of colours used for the large range in Experiment 3 was not as large as that in Experiment 1, but still contained good examples of green and blue colours; this would appear necessary, though not sufficient, to use the green–blue boundary as the naming boundary. It was not sufficient because the prior exposure condition of Experiment 3 showed that a bias could be introduced that would affect the position of the green–blue boundary even if good examples were given in the naming task. It is also clear that given a restricted range without clear examples (i.e., close to prototype) of both blue and green, naming like MTS will tend to divide the stimuli into two equal groups.

**GENERAL DISCUSSION**

Colour boundaries are assumed to be stable, and hence one might assume that categorical similarity would be used for all tasks where colours need to be distinguished. Our main finding is that, despite the efficiency gained by allocating items to a category (see Smith & Medin, 1981, for discussion of this normal advantage of having acquired categories), categorical similarity is not used in a large number of conditions. We note here a similar conclusion in the auditory domain for phoneme categorisation (Gerrits & Schouten, 2004). It was only in the visual search task (Experiment 3) where we found that performance was based on existing colour boundaries.

What then is the mechanism underlying CP? We propose that the default position in processing colour similarity relies on a neuronal network that automatically associates language areas and visual cortex (Siok et al., 2009) with the corollary that colour similarity will be categorical. The default judgements about colour are a result of an underlying representation that has a symbolic structure but with the important phenomenal consequence that colours from different categories look more different than would be predicted from our ability to discriminate between them. Thus, we provide an explanation for the everyday observation that rainbows look striped even though the colour change is continuous. The default position would apply to most judgements concerning colour that implicitly ask for an assessment of colour similarity. We note what might be a connected proposal made recently by Filoteo, Lauritzen, and Maddox (2010) for category learning. Filoteo et al. surprisingly found that it was in implicit, rather than explicit, category learning that there was an increased use of optimal strategies.

An alternative proposal for the origin of CP (Roberson et al., 2009) would account for our data by predicting a conflict between verbal and perceptual codes. In their proposal, both verbal and perceptual codes are automatically activated and come into conflict only for within-category trials when the verbal code gives a “same” response for the colours but the perceptual code responds that they are “different”. The Roberson et al. (2009) account may have the added benefit of explaining why the assumed cross-category pair in Experiment 3 produced no advantage for the green-shift groups. If observers are inclined, through their recent experience, to call/see both stimuli of the cross-category pair as “green”, then that pair would produce a similar response to all the other pairs. However, the conflict between automatic activation of both verbal and perceptual codes would not predict our MTS data.

We now turn to consider conditions where the categorical judgement of colours in implicit tasks can be disrupted (cf. the disruption of the automatic task of reading; Stolz & Besner, 1999). One condition is the addition of a concurrent language task (Gilbert et al., 2006; Roberson & Davidoff, 2000; Winawer et al., 2007). Indeed, in these interference studies, observers only have to remember a single colour or detect a single odd-one-out colour; so, the categorical similarity procedure must be easy to disrupt. What the present paper shows is that there are procedures in explicit tasks, and ones that are not obviously demanding of language resources, that can also prevent the use of colour categories. Our data for MTS certainly support that conclusion as presumably it is sufficient to promote perceptual similarity by making the observer carry out some near identity similarity judgements. The deactivation of the default colour category networks in an explicit colour naming task might seem more surprising. It would be reasonable to believe that
being asked to name a colour patch green or blue would inevitably activate the existing colour boundary rather than do the exact opposite as it did for the green-shift group of Experiment 3. Instead, the data from Experiments 1 and 3 suggest that the requirement for the use of colour boundaries in colour naming is the presence of good examples of the colour in the samples to be named.

Once categorical similarity has been deactivated in MTS, responses will be affected by the range (Experiment 1) and by the perceptual salience of a focal colour in the range (Experiment 2). Thus, the types of cognitive processes activated are those that one might expect to operate in any categorisation judgement based on perceptual similarity (for reviews see Hampton et al., 2005; Murphy, 2002). In MTS, one in particular that would appear relevant is the positioning of any decision criterion in relation to the prevailing range of stimuli by moving the criterion towards a running average of the range (Treisman & Williams, 1984). Nevertheless, one must ask why we should want to be prevented access to categories already mastered. Tanaka et al. (1998) addressed a similar question of when a categorical or noncategorical strategy is used in a face MTS task. Referring to work on speech sounds, Tanaka et al. argued that categorical judgements are used only when there is an automatic use of category structure (Kuhl, 1991). With respect to CP, our data largely support that argument.

In summary, several processes operate to make a colour similarity judgement based on perceptual similarity rather than the default categorical similarity. For language impaired patients, as Goldstein (1948) noted, perceptual similarity is all that is available to make colour decisions. In consequence, the outcomes of their colour sorting are bizarre. Normal observers may also use perceptual similarity if that is what instructions demand (Kay & Kempton, 1984). However, normal observers may apply other constraints to override perceptual judgements. One obvious constraint is to use colour names. Nevertheless, in MTS, even with good examples of green and blue as samples, colour names are not applied by observers in our studies. Instead, what is brought to bear is a decision criterion based on perceptual similarity with the important consequence of categories based on a binary division.

Original manuscript received August 2011
Revised manuscript received June 2012
First published online July 2012

REFERENCES


